

## DESCRIBING THE INTERACTIVE DOMESTIC ROBOT SETUP FOR THE SERA PROJECT

**Sarah Creer<sup>1</sup>, Stuart Cunningham<sup>2</sup>, Mark Hawley<sup>1</sup>, and Peter Wallis<sup>3</sup>**

<sup>1</sup>*School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom*

<sup>2</sup>*Department of Human Communication Sciences, University of Sheffield, Sheffield, United Kingdom*

<sup>3</sup>*Department of Computer Science, University of Sheffield, Sheffield, United Kingdom*

□ *The Social Engagement with Robots and Agents (SERA) project conducts research into making robots and agents more sociable. A robot setup was deployed for ten days in the homes of users to generate audio-visual data for analysis on the nature of the evolving human-robot relationship. This paper details the setup developed to provide opportunities for human-robot interaction and yield the quantity of data required for analysis. The robot's function was not to exist as part of an experiment but to exist in the user's home, fulfilling a role in his/her existing routine to ensure interaction. The system acted as an exercise monitor to encourage older people to lead a healthy lifestyle. The assumption made was that increased engagement and usefulness of the system leads to increased use, providing more data for analysis. This paper describes the SERA robot setup for each of three iterations of deployment, with particular reference to maximizing the amount of data collected.*

### INTRODUCTION

Robots are already appearing in homes in the form of toys; for example, Robosapien (WowWee); as appliances such as the Roomba vacuum cleaner (iRobot); and as wholly new applications such as the Nabaztag (Violet). Considerable research funding has been directed toward exploring the potential for integrating other technologies into such devices. The vision is often based on the robot-as-butler metaphor, and the possibility is raised of a spoken language interface to some form of domestic or work-based robot. Dictation systems and telephone-based interactive voice response (IVR) systems are available for daily use in the home, but more knowledge

The research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007–2013] under grant agreement no. 231868. It was also supported by Loquendo, with access to their TTS product.

Address correspondence to Sarah Creer, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK. E-mail: s.creer@sheffield.ac.uk

is required about the use of far-field microphones in a user's home environment for less structured, more natural interactions.

The question of to what extent a robot in someone's home setting would be treated as a social actor is unknown, along with what strategies are available for dealing with notions of role, politeness, and power relations. On the path to addressing these issues, the SERA project set out to collect audio-visual data of a real robot interacting in a real domestic environment.

The interactive robot setup was deployed in participants' homes for 10-day periods three times throughout the project. For each deployment, a different set of behaviors or functions was implemented. The aim was not to empirically test hypotheses using the iterative changes or to develop a perfect system or application but to collect data that could provide new insights into the evolving human-robot relationship corresponding to each set of behaviors.

This paper describes the setup of the SERA project robot for each of three iterations, highlighting the aim of maximizing data collection for subsequent analysis, and providing background information and context for the data-analysis-based papers in this issue.

## OVERVIEW

Deviating from a more traditional laboratory-experiment-based setting, which depicts a short-term, usually one-off task or goal-based interaction, the SERA project collected audio-visual data to inform on the nature of the robot's involvement with the participants during the course of their normal daily lives over a long-term period. The 10-day trial period allows enough time for the novelty factor of the robot's presence to wear off and long-term engagement to be studied.

With each deployment of the robot, additional new participants were recruited and those involved in previous iterations were retained. The aim was to study the relationships both over the trial period and across the three deployment periods where possible.

The iterative nature of the project allowed the alteration of the behavior or functioning of the robot for each of the three periods of deployment, based on introducing procedures to address situational or contextual needs, using theory from relevant literature, making observations from both the video data and post-trial participant interview data, and following the research interests of the consortium.

Using a small sample size of up to three new participants per iteration, the intention was not to set up experiments and test hypotheses nor to try and move toward an ideal system or application, but to implement different sets of behaviors for each iteration and to collect data that could more

widely inform research on social interaction and relationships among humans and robots.

The assumption made throughout the project has been that increased sociability of the system leads to increased use of and engagement with the system and therefore to more data being collected for analysis. This assumption is based on the findings of, for example, Krämer, Bente, and Piesk (2003), who found that using embodied agents as interfaces provoked socially desirable behavior from the humans and increased user expectations of the agents' social capabilities. The underlying objective while making changes to the setup was therefore to identify and alter those features that were seen to be potential obstacles to sociability and to hypothesize which new features could improve sociability in the robot. The initial setup and these changes had also to fit within the limitations of the technology in use.

## **SETUP**

The SERA project used a stand-alone robot setup, which is always switched on and which has the potential to interact at any point, resulting in it being a continual social presence in the home. Its embodiment practically places it in a specific location where, because the participant has an expectation that he/she can be heard and seen by the robot, interaction is possible and they can attend to each other.

The robot's presence and embodiment as an entity in the home accounts for its ability to actively initiate conversation as opposed to a traditional computer interface which acts as a passive responder to commands. This is particularly relevant in those instances when increasing interaction results in more data for analysis.

The following section details the components of the SERA setup which was used throughout the three iterations. It is shown in Figure 1 with a schematic diagram of the hardware components in Figure 2. Three systems were available for deployment. This restriction on hardware constrained the number of participants that could participate in the trial within the defined time period for data collection. During iterations 2 and 3, two systems were used for deployment, with the third as a system to continue development throughout the data collection time period and to ensure a fast resolution to any technical problems by substitution of hardware or debugging of software issues.

## **Application**

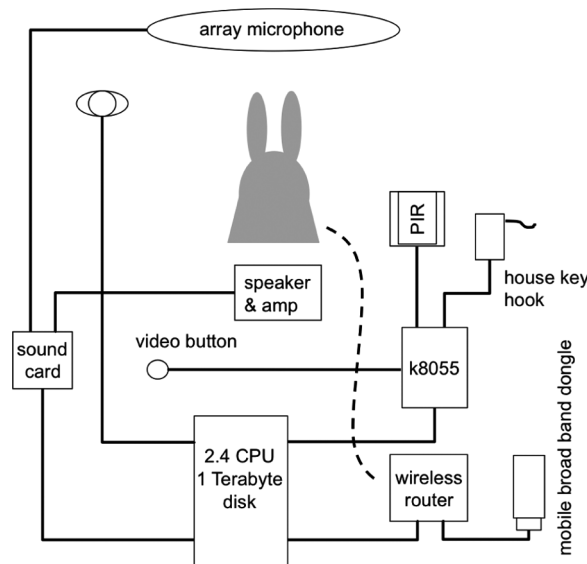
To encourage use of the system to gather the data for analysis, the robot needed to be seen to have a purpose and a context within which to interact.



**FIGURE 1** The setup showing the Nabaztag and the stand housing the component parts placed in the hallway of a user's home.

To maximize collection of data, the application provides a useful reason for interaction and a routine within which the interactions could occur. The system was designed to be part of a real, current, and useful application context in that it could also contribute to research in the field of assistive technology for older people. It aims to contribute to management of the increased demand on health systems for care for older people that is predicted by the demographic change existing in the developing world. The application is a system that aims to prolong independent living by helping older people maintain their health through living an active lifestyle. Assistive technology applications allow users to become more involved in taking responsibility for their own health, contributing to the change of behavior that may be required to help manage existing conditions; and they have the potential to positively affect users' health, well-being, and independence (e.g., DelliFraine and Dansky 2008; Paré, Jaana and Sicotte 2007).

Specifically, the SERA application was developed to assist older users to adopt and maintain an exercise routine over time. The development of the application used the Transtheoretical Model of Behavior Change



**FIGURE 2** Schematic diagram of the hardware architecture.

(Prochaska and Velicer 1997) to target users. It states that there are five stages in which a person can be when changing his or her behavior in some way: pre-contemplation, contemplation, preparation, action, and maintenance. The application initially targeted those who were in the preparation and action phases of changing their behavior to do more exercise. Throughout the data collection, the application used planning and goal-setting, creating awareness through self-reflection and self-monitoring of activity, and building self-confidence in the user's ability to lead an active lifestyle.

Personalizing the system to each individual's needs, the users provided the system with a self-devised activity plan, which contained details of what exercise the user had planned throughout the 10-day trial period. The activity plan, or diary, therefore became a knowledge base of information from which a more specific context of discussion could be taken.

Previously developed applications such as the FitTrack system have combined health-related, task-oriented interaction with building long-term social bonds between a human and an embodied agent (Bickmore, Gruber, and Picard 2005). That system was designed purely as a computer application used to log and discuss exercise rather than a social presence that could observe the participant and his or her behavior. The presence of the robot setup contributes to the persuasive aspect of the technology, reminding the user of its existence and therefore its function, in this case, monitoring and encouraging exercise. The SERA system provides an opportunity to investigate the differences between traditional passive

interfaces and embodied pro-active interfaces and to determine the differences that a more sociable tool makes in terms of adoption and acceptance of technology and persuasion to change behavior for health benefits.

## **Sensors**

The initial design of the setup used smart-home technology to obtain knowledge of where the occupant was in the house and any activity the user performed throughout the day, with the purpose of using this knowledge to inform the interactions. The usual approach is to put sensors on doors, in rooms, and on home appliances in order to provide detailed information on the behavior, activity, and daily routines of a participant. The disruption made by the installation of sensors or video cameras throughout people's homes, however, provides an obstacle to recruitment of such projects and acceptance of the technology. Where there is no immediate benefit to the participant or health need, such as emergency alerts, the monitoring of overall activity and behavior can be intrusive and therefore an alternative was sought.

For the SERA project, the most basic information required was the proximity of the user to the robot and information relevant to the context of a health and exercise-based application. A knowledgeable "smart area" was constructed, consolidating the additional information sensors into one portable and easily installed setup containing the hardware components and a stand for the static robot. A passive infrared sensor (PIR) was placed in the stand; this indicated the close proximity of the user, therefore ensuring that the system would provide output only when the user was close enough to attend to it.

In addition to the proximity sensor, a key hook switch was added to the Nabaztag's stand where the users were asked to store their house keys (or an equivalent weight if users were unable or unwilling to store their keys on the stand). The key hook switch approximates fixing sensors on the outside door to indicate the person going in and out of the house. It initiates interaction based on going out to do some outdoor activity and coming home after the activity has been done. In combination with the key hook switch information, the user's self-devised plan of activity and the system clock were used together as a knowledge base that provided information about what the participant was likely to be doing throughout the day, relevant to his/her exercise and activity.

## **System**

A modified Nabaztag was used as the embodied robot for the SERA project and therefore as the intended focus of the system. The robot takes the

form of a stylized rabbit with movable ears and flashing colored lights on its front. This robot was selected because researchers involved in the project had knowledge and experience using the Nabaztag and understanding of the overlaps with other simultaneously running projects at the institutions within the consortium.

At the system's core is a low-power and minimal-noise computer (VeryPC). The software is primarily written in Java, with a socket connection to a k8055 hardware interface board (Velleman) which links to the sensors and can take input from the users via buttons. A video button is connected to a Webcam Pro 9000 (Logitech) which determines the storage of the video data as controlled by the user. The connection to the robot is via a wireless router that also connects to the internet via a pay-as-you-go broadband dongle (3). The computer has an external sound card connecting it to the VoiceTracker array microphone (Acoustic Magic) and to the external speaker mounted below the robot via an amplifier. The aim was to produce a robot that could use speech recognition as input from the user. This interface was not implemented within the lifespan of the project and alternative interfaces were used.

## **Dialogue Management**

The dialogue manager was a state-based system written in Java. The system entered a state on initiation of a dialogue and from that point was always in a state setting the context of the discussion. It used pattern-action rules that determined what actions were taken dependent on input from the user and any other set conditions. The actions that could be taken were: producing output, updating actions within the state, or movement to another state.

The dialogue state networks were developed to be self-contained and closed, in that there had to be a path for every possible input from the user. This included there being no response from the user or there being an unexpected response taking the user away from the current topic. The management for these situations is expanded in the relevant dialogue sections in the iteration set-up descriptions. The aim was to develop coherent dialogue that would follow as far as possible the intuitive strategies employed by humans in human-human interaction, for example, providing appropriate ways to open and close conversations and continuing the dialogue, appropriately taking into account a conversation partner's response (see Heylen et al. [this issue] for further discussion on this topic).

The networks expanded through the iterations from 23 different states, where each state represents a spoken turn from the robot that is followed by a transition to the next turn, dependent on the response from the user,





recruited for the project is shown in Table 1 together with their demographic details and in which iterations they were involved.

The participants were visited by one of the researchers who explained the project in more detail and collected their consent to participate when appropriate. Before the trial began, the participants provided the researcher with a self-devised activity plan for the trial period, which was then input into the system. The activity plan detailed their planned activity for the period including any other diary events that they were happy to provide, and it could form the basis of an interaction with the robot.

The robot was installed in the users' homes in a position that pleased the participants and that maximized privacy for them, the other people who lived in the house, and any visitors. The initial site suggested was in the participant's hallway to allow reasonable function of the setup as storage for the participants' house keys; it was felt to be the least intrusive position, ensuring frequent passage but less sustained presence. The priority on positioning the system was placed on the participant's preference. On installation of the robot, the system was explained to the participant in terms of the interaction method and what types of interactions the system would have with them. The participants were given a booklet of instructions, in case they needed some clarification of how the system worked, together with all the contact details for the researchers. The participants were contacted after three days to ensure that they were happy to continue with the trial and also to make sure that the technology was functioning as expected.

From that point until the end of the trial, the participants were not contacted by the researchers unless they initiated contact with questions, problems, or organizational issues. There was no remote access to the rabbit, except for sending messages to the robot, and therefore no real time viewing or interference in the function of the system. In addition to having videoed data for analysis, the participants were left with a notebook to use if they wished to make notes on what they liked or disliked about the robot or the trial itself. The notes were then used to help prompt questions during a follow-up interview, which took place after the trial had finished.

**TABLE 1** Participant Information: Participant Number, Sex, Age, Whether They Live on Their Own or with Others and in Which Iterations They Were Involved

Participant	Sex	Age	Live alone?	Iterations
P1	F	65	Y	1, 2, 3
P2	F	50+	N (1 other)	1, 2, 3
P3	F	60	Y	1, 2, 3
P4	F	50+	N (1 other)	2, 3
P5	M	76	N (1 other)	2, 3
P6	M	71	N (1 other)	3

## ITERATION 1

The purpose of the first iteration was to provide a test of the approach to data collection and to determine what factors were essential for a successful deployment, getting quantity of data and what could threaten acceptance of the technology and participation in the study.

A second purpose was to provide a baseline of the interactional quality that the subjects provide to compare with the subsequent iterations of data collection. It was an opportunity to test the technology components for following iterations which have an increasing number of participants and increased complexity in the setup.

Finally, iteration 1 would serve as a test for the data analysis methods being used and would demonstrate how the researchers would use the data and what would be required in order to be able to pursue those research lines.

Three participants were recruited for iteration 1: P1, P2, and P3 (see Table 1).

## Method of Interaction

The original aim was to implement automatic speech recognition (ASR) into the setup to take input from the user. Using off-the-shelf ASR in the system proved to be unsuitable for the dialogue application and therefore implementation was postponed for the deployment. The alternative was to mount “yes” and “no” buttons on the robot’s stand to take input from the participants. Output from the robot was provided using synthesized speech of the female UK voice from *Dragon Naturally Speaking* preferred version 10 (Nuance).

## Dialogue Content

The dialogue content was an implementation of an exercise monitor and self-reflection program from the British Heart Foundation publication, the Heart Failure Plan (Lewin, Pattenden, Ferguson, and Roberts 2005), adapted to fit with the interaction limitations of the technology. It states that exercising without a plan can lead to the danger of doing too much exercise on one day followed by a need to rest on subsequent days and difficulty in doing enough exercise from that point onward due to being overtired from the original activity. Following that pattern of behavior leads to the “over activity-rest” cycle ultimately leading to a loss in fitness and motivation to continue building a healthy lifestyle. To avoid the cycle, the recommendation is to build a plan of activity that contains manageable amounts for that particular individual, to be done regularly. It is fitted to

the abilities of the individual through altering the amount or type of exercise based on self-awareness and reflection on how he/she is feeling after adhering to his/her plan.

Aiming to maximize data collection, initiation of interactions could be done not only by the user but also by making the robot an active initiator by using the sensor information. The initiation events were:

- first appearance of the participant in the morning (activated by the PIR sensor and using the system clock)
- the participant goes out (activated by the key hook switch and using the system clock and diary information)
- the participant returns home (activated by the key hook switch and using the system clock and diary information)
- after the last activity of the day (activated by the PIR sensor and using the system clock)
- the system receives a message (activated by the PIR sensor and message detection)
- user-initiated message retrieval (activated by the PIR sensor and button press)

### ***Overall Dialogue Guidelines***

The aim was for the dialogues to be natural sounding, polite, and to produce relevant and clear content, to minimize annoyance for the user and to allow the user to opt out of the interaction. These were factors considered to contribute to the participants' continued use of the system throughout the trial period.

The language avoids using "I" and uses mostly the passive tense. This follows Nass and Brave (2005) who claim that using "I" should be avoided in systems with yes/no buttons as it can be perceived as controlling or unfair to use speech and display individuality when the users themselves cannot do the same. The language is more introvert than extrovert except where declarative sentences are used during exercise-related interaction.

### ***First Appearance in the Morning***

The *good morning* dialogue was initiated by activation of the PIR sensor and the knowledge that this dialogue had not been initiated that day. The robot greeted the user and asked if they would like to hear a weather forecast. Hearing a weather report allows the user to reschedule his/her exercise for the day dependent on the outdoor conditions. The participant was then asked whether he/she had weighed his/herself. This information is not essential for tracking fitness but it is potentially useful for future assistive technology applications where it can monitor the behavior of

individuals and potentially provide alerts for care for those with long-term conditions. The participants were told in advance that this would be one of the features of the robot application and they were not obliged to record their weight or provide that information as part of the video data.

### ***Going Out of the House***

The *going out* interaction was initiated by taking the house keys from the hook on the stand. The activity diary is checked and if there is a specified activity at that time, then it is referred to in the output. If there is nothing specified in the diary a more generic output is produced.

### ***Coming Home***

The *coming home* interaction was initiated by replacing the keys on the hook on the stand. The activity diary is checked and if there is a specified activity for that time then questions are asked dependent on that activity and whether the activity plan has been followed. The importance of adherence to the activity plan is reiterated and encouragement to do so is provided.

### ***Self-Reflection after the Last Activity of the Day***

The *self-reflection and evaluation* interaction was initiated by the PIR sensor during a specified time period after the completion of the last activity of the day. The participant was asked how he/she was feeling and was asked to evaluate whether the amount of activity he/she had done that day was suitable. If the user decided that it was, then he/she is encouraged to follow the activity plan again, and if he/she decided that it was too much exercise, advice is offered on taking breaks and reducing the amount of exercise. The main function of this interaction is to offer the user an opportunity to reflect on the exercise that has been done and relate that to his/her well-being.

### ***Receiving a Message***

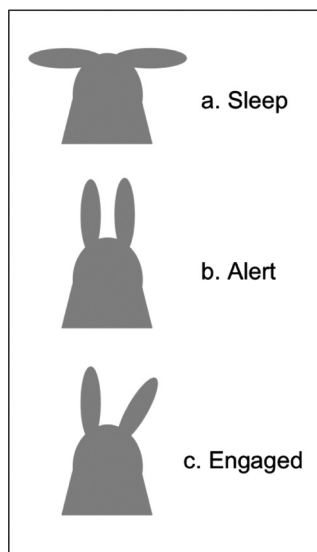
The *message* interaction was initiated by detecting a message input to a Web page in combination with PIR sensor activation. The message was personalized for the user and was either from the researchers or from a non-explicit source, it was never stated that the message was explicitly from the robot. The messages were of three types: information about the activity plan, information about the system or project, and information such as entertainment listings or social activities that may interest the user. The aim was to send one message to each user per day.

### *User-Initiated Message Retrieval*

On approaching the system and pressing the “yes” button, the user was offered a chance to hear an unheard or previously heard message, and after that he/she was offered the opportunity to hear an up-to-date weather report or information about the system.

### **Attention Management**

Due to its continual presence in the home, the robot needed to have information about when the user is nearby and whether he/she is willing to engage with it. For iteration 1, this was defined in association with the dialogue management and the initiation events and conveyed to the user by its ear positions, shown in Figure 4. The ears were positioned horizontally as the default or sleep position (a. in Figure 4). At the beginning of every interaction, the rabbit’s ears rotated to the alert position (b. in Figure 4), pointing upward. The ears returned to the sleep position once the recording has stopped (approximately one minute after the initiation). On activation of the PIR sensor, the rabbit’s ears rotated to the alert position and if there was no interaction they returned to the default sleep position. When the user-initiated message dialogue was activated, the rabbit moved one ear slightly forward into the alert position (c. in Figure 4). Eimler, Krämer, and von der Pütten’s study (this issue) investigated the relationship between the ear positions displayed by the robot and the



**FIGURE 4** Ear positions used by the Nabaztag to indicate its state of attention.

**TABLE 2** Showing the Number of Videos Gathered for Each Participant During Iteration 1 of Data Collection

Participant	N videos
1	35
2	10
3	7
<b>total</b>	<b>52</b>

associated perception by humans, informing the use of ear configurations to convey attention status of the robot.

**Data**

The amount of data collected during iteration 1 is detailed in Table 2. Each video lasts between approximately 30 seconds and 2 minutes and shows an interaction, as described above, between the participant and the robot.

**ITERATION 2**

For the second iteration, various features of the setup were changed to provide more relevant functionality for the participants, overcome various practical obstacles, and present a more engaging interaction hypothesized to result in more interaction and therefore more data.

P1, P2, and P3 were retained for iteration 2 and two new participants were recruited: P4 and P5 (see Table 1).

**Method of Interaction**

An alternative method of interaction was provided to allow for more varied and expressive user input and for more variation in the robot output and functionality. It was an attempt to progress toward the more natural speech interface by providing more expression and variation. The need for this change is reinforced by the results of analysis from the iteration 1 interview analysis (Klamer, Ben Allouch, and Heylen 2010; Klamer and Ben Allouch 2010a, 2010b). The method of interaction had to involve direct engagement with the robot without any interference of another feedback device, such as looking at a screen when using a keyboard. The planned speech recognition component was not yet integrated to the system and therefore the in-built Radio Frequency Identification (RFID) tag reader in the Nabaztag was used to allow the robot to read words and symbols on tagged cards. The set of words and symbols represented the

different topics of which the robot had some knowledge and interaction items that would allow the user to input appropriate responses to the exercise-related conversations. This method of interaction allowed more user-led interaction, including more user-initiation.

The tags were placed up against the front side of the robot, below its facial features. The lights on the robot's front flashed to indicate that the RFID tag had been successfully read.

The application was extended to allow the user to input durations of exercise in minutes that they had done in a day, switch between topics of conversation (which had previously been dealt with using yes/no questions) and rate how they were feeling after a day's exercise. The cards used to perform these tasks are listed in Table 3.

For the first iteration, the yes/no interaction meant that at every turn the robot was in control of the conversation, demanding responses from the user. For the second iteration, the new interaction method could take advantage of not requiring specific responses to move through the interaction. As is found in natural speech, the content of a turn is not necessarily important but the fact that it exists maintains the interaction. Tokens such as "mm" and other backchannels can further the conversation as ongoing, optionally inserted acknowledgements.

The design choice of the set of cards in Table 3 of using symbols for the "interaction" cards meant that these could be multifunctional in their meaning. A smiling face could indicate "yes," or act as an acknowledgment, a backchannel, an agreement, or an expression of a positive feeling. A

**TABLE 3** RFID Tagged Cards Used for Iteration 2

Interaction set	Smiling face
	Neutral face
	Frowning face
	REPEAT
Topic set	WEATHER
	SUMMARY
	ADD TO LOG
	SYSTEM
Exercise duration set (in minutes)	MESSAGE
	10
	20
	30
	40
	50
Rating set	60
	1
	2
	3
	4
	5

frowning face could indicate “no,” a disagreement, a suggestion that the system was wrong in what it was saying, or an expression of a negative feeling. Using pictures instead of text may provide variability in the interaction, but with that, introduces the need for interpretation of what the user is trying to convey. For example, when told how much exercise is in the user’s diary for that day, showing the frowning face could either indicate that the system has given an unexpected or incorrect amount or the user could be indicating that he/she is not happy to be doing so much exercise. The dialogues have to be able to interpret the input correctly or at least recognize when there is ambiguity in the response. With every increase of flexibility for the user, the dialogue becomes less constrained and the possible path to follow in an interaction becomes more complex. Keeping the dialogue constrained and comprehensive, while allowing more conversational power to the user is a continuing challenge.

### Dialogue Content

The expansion of the possible input vocabulary expanded the functionality to continue to assist with self-monitoring and self-management of an exercise routine. Observations made from the data resulted in the view that the users were generally not in the targeted preparation or action phases of the transtheoretical model, but were in the *maintenance* phase, already successfully following an activity plan. Maintaining behavior change requires a different set of strategies for the application, and the application required some modification. Further details of this analysis can be found in Wallis, Maier, Creer, and Cunningham (2010).

To maintain a high level of use of the system and to maximize data collection, the dialogue content had to be made more relevant to the existing participants and also to the new participants recruited for further iterations. The dialogue content therefore focused on building self-efficacy. Self-efficacy has been defined as the context-specific confidence that an individual has in their ability to fulfill their goals (Bandura 1977). In this case, the goals were defined as a daily amount of planned exercise as stated in their activity plan, contributing to a more general aim of living an active, healthy lifestyle.

Building and maintaining high self-efficacy helps keep users active in changing their behavior and helps them to not relapse into their previous pattern of behavior; this makes it relevant for users in all stages of the transtheoretical model. The system builds self-efficacy by reminding the users of what exercise they have planned to do throughout the day and keeping track of their progress toward this goal. Input was taken about how much exercise the participants completed, which was subsequently fed back to them at the end of the day and was compared to their daily exercise goals.



In addition, self-evaluation is performed, allowing time for the participants to reflect on how they are feeling and to become aware of the effect that their exercise routine is having on their well-being. Having a record of what exercises they are doing not only provides the users with an opportunity to motivate themselves and to see when they are succeeding, but it also provides realistic information about their previously estimated behavior. Providing information about performance is a vital prerequisite to self-motivation and taking action. As the daily goals are set and reached, they can provide markers of achievement that increase over time to increase the user's self-efficacy and therefore maintain a level of exercise (Bandura 1998).

The initiation events are the same as for the first iteration, however, much more use is made of the user-initiated dialogue as described below. The number of states involved in the overall dialogue structure has increased to approximately three times that developed for the first iteration, introducing a wider variability in the range of output heard by the user.

### ***Overall Dialogue Guidelines***

During the interactions, the robot could be asked to repeat itself. In an attempt to repeat the turn as a human might do, the robot first responded with an almost direct repetition of the turn prefixed by a hedge, for example, "I just said..." For a first repetition, the user was assumed not to have heard the output properly. For a second time, it was assumed that the user had not fully understood the turn, therefore the prefix was "I was just saying..." and the rest of the output was rephrased in some way. This clarified the content meaning and restructured it in case of confusion caused by the speech synthesis output.

Following from analysis of data centered around closing rituals from the first iteration, changes were made to the endings of the dialogues to make them more explicit, while still encouraging the user to continue the conversation if he/she so wished. At the end of the dialogue, the robot yielded control to the user, suggesting an optional end to the conversation: "Unless there's anything else you'd like to talk about, I'll talk to you later." The analysis of the dialogue endings relating to this issue is found in Payr (2010).

The language used by the robot was changed from the first iteration to use self-reference due to the expanded and more expressive input allowed for the user. The development of a persona for the robot, which provided input into the language used, is discussed in more detail in *Persona*.

### ***First Appearance in the Morning***

As in the first iteration, at PIR sensor activation a greeting is output. If the greeting is responded to using a card, the conversation continues. If

there is no response, then the conversation is halted and a timer is set to 10 minutes. During this period, if the PIR sensor is activated, then the dialogue is not initiated. After this time period, on PIR sensor activation, the procedure is restarted and continued until the greeting is responded to. This procedure ensures that the greeting is delivered to the user rather than anyone else in the house who may have activated the sensor. It also allows the user to ignore the robot until they feel ready to have a conversation. Giving the user this control over when the conversation occurs, while being reminded that it should occur by the continually activated greeting, also addresses the need to provide opportunities for interactions to maximize data collection.

In the second iteration, the robot phrased the weighing question as follows: "Some people find it useful to weigh themselves daily as part of their exercise routine. I hope you don't mind me asking, but will you be weighing yourself today?" In contrast to the previous iteration's more direct question, it provided a reason for asking the question, recognized and acknowledged that asking it is a potentially face threatening act, a speech act that may oppose the desires of the conversation partner (Brown and Levinson 1987), which in turn appeased the act itself.

The users were generally aware of what exercise they had planned. However, to validate their exercise schedule, the users were provided with their goals for the day, taken from their activity plans, and presented as a total number of minutes. Providing more information for the user and variability in the interaction, the dialogue content depended on what exercise the user had planned for the day. For over 30 minutes of exercise, National Health Service (NHS) recommendations are stated, specifically that a healthy adult should be doing 30 minutes of exercise, five times a week to maintain a healthy lifestyle. This was an attempt to acknowledge and reassure the user of the right course of action required to lead a healthy lifestyle. If the user had some exercise planned but totalled less than 30 minutes, then the NHS recommendation is held back as this would be providing negative feedback and could detract from the self-efficacy that is being built. If the user has no exercise planned or the amount planned is less than 30 minutes, then there is a reminder that exercise can be added to the exercise log, and it is reasserted that activity such as housework, gardening, or walking can be added to the log. This builds awareness of all the possible ways that exercise can be incorporated into a healthy lifestyle rather than just focusing on explicitly planned exercise, such as exercise classes.

Basing the dialogue content on conditions and taking information from the activity plan allowed the dialogue to be variable with respect to the planned goals and abilities of the user, compensating in part for the inability to alter the users' goals within the system during the trial period.

For the users to be aware of the goal that they have set for that day and have the reinforcement depending on their plans, the morning conversation needs to occur in full. If the user tries to change the topic of conversation before all the information has been imparted, the robot does not allow the change, but asks the user to allow the robot to finish. The idea of *mixed initiative* interactions, where the agent would allow a change of topic and, once that deviation is complete, remember the discussion that was being had and then return to the previous interaction to continue the conversation, means that the robot would be able to yield control to the user with respect to the flow of conversation. The current organization of the state-based dialogue manager does not allow this type of behavior and, therefore, to ensure that the health application functions as designed, with the reassertion of the goal of activity planned for the day, the robot asks the user to remind them to change the topic at the end of the dialogue. This may introduce points at which the user has to yield the floor and the control to the robot, although it is the only way in which the full amount of information with regard to the health-related interactions will be passed on to the user.

### ***Going Out of the House***

Similarly to the first iteration, when the keys were removed from the hook, the diary was checked for scheduled activity. If an activity was in the diary, it was stated in the dialogue output, showing that the robot has stored that information.

The robot acknowledged the exercise and reminded the user to add the amount of minutes done to the log on their return. These dialogues, expanded from the first iteration, allowed the user to tell the robot that it is using wrong information by showing the robot the frowning face. The user was asked to confirm this as the correct interpretation of that card, and the information can be flagged as incorrect, which can then influence the direction of the coming home dialogue. With no activity recorded in the plan, the user was asked if he/she was going out. If the user responded “no” with the frown card, then a timer was set so that on replacement of the keys the coming home interaction was not activated. This procedure was introduced to allow participants to indicate to the system that they were using their keys for purposes other than leaving the house, to open the door to visitors or take their rubbish out to the bins, for example.

### ***Coming Home***

On returning home and replacing the keys on the hook switch, if the users have not been doing any exercise (as indicated in the system diary)

they were asked if they had a good time, to provide some social politeness. The system asked whether anything should be added to the exercise log, promoting self-awareness of and reflection on what the users had been doing, and its contribution to the exercise log. This procedure aimed to make users reflect on what they might not have thought of as being exercise but could be added to their log, such as carrying shopping bags. This is dependent on the user's interpretation of exercise and what he/she wanted to store in his/her log.

If the user had been doing some scheduled exercise, then the robot asked if he/she had a good time and used the information from the diary to suggest the amount of activity that should be added into the log. This not only reminds the user of the daily goal but also affirms that it has been achieved or exceeded, contributing to building the user's self-efficacy. If the user does not achieve the planned exercise goal, awareness is created, but with no rebuke from the robot that could be detrimental to the user's self-efficacy level. The robot summarized the amount of exercise done so far that day, indicating that the input had been registered and providing positive feedback and reinforcement.

If, on the way out of the house, the user indicated to the robot that the activity taken from the diary is wrong, then the robot uses the dialogue that would be activated if the user was coming back after an unknown activity.

#### *After the Last Activity of the Day*

After 5:00 p.m. and indicated by the PIR sensor, the user was invited to become engaged in a conversation with the robot about how they are feeling and how much activity has been done in comparison with what had been planned.

The user was asked if he/she was available to talk, allowing the user to be in control of when he/she would interact with the system, which used the same timed initiation mechanism as the morning dialogue.

If information is available from the same conversation from the previous day, this is fed back to the user as a reminder of what exercise was done yesterday and how he/she felt (on a rating score of 1–5). This was used to calibrate how the users were feeling and to relate the amount of exercise to the rating. After feedback from the previous day, the users were provided with information about that day's exercise. If they had exceeded or matched their planned activity goal, they were told both their goal and a summary of how many minutes of exercise they had done. If their goal was not achieved, they were not reminded of it, but they were provided with the number of minutes of exercise that they had done. If there were zero minutes of exercise recorded for that day, there was no rebuke or reminder of the original goal. This approach was made to ensure that there was no

negativity about the amount of exercise done and that self-efficacy was maximized. Once the amount of exercise had been provided, the users were asked how they were feeling and to rate that feeling. This two-stage process attempted to ascertain the users' expressions using the smile/frown/neutral faces and potentially additional verbal responses followed by a concrete, unambiguous rating (1–5) that was stored for use in the following day's equivalent discussion. This opportunity for the users to validate their own activity levels and increase their awareness of their physical and mental well-being aimed to increase confidence in their abilities and motivation to follow their activity plans.

### *Message from the Researchers*

Using the message system devised in iteration 1, the system checked for a message on a Web page at frequent time periods and, if there was a new message, at the next activation of the PIR sensor the message dialogue was initiated. The user was asked if he/she was ready to hear the message and, if there was a positive response, the message was output. With a negative response, the message was stored for future access. This again allowed the robot to aim the message at the participant rather than any other member of the household and allowed the user to control when the message was heard. It was aimed to provide one message per day for each participant.

### *User-Initiated Dialogue*

Expanding the user-initiated conversations provided more opportunity for data collection and promoted use of the system with the user in control. It was initiated by the user showing the robot one of the face cards. The robot responded with, "Hello?" and waited for the user to continue the conversation by selecting a topic or another card. If any other card was used, then the robot asked how the user was and made suggestions about how the conversation could continue. The topics available to the user were: adding to the log (also initiated by the 10–60 cards), finding out the weather, providing a summary of exercise done (yesterday, today, or in total), or finding out more information about the system. These topics were available to access at most other points in the interactions. Adding to the log was done through this initiation to account for any exercise done without indication from the key hook switch. Finding out a summary of the exercise done so far allowed constant access to the rising total, used in comparison with the previous day's exercise; or, if the participant had not done exercise that day, then his/her efforts could be validated by accessing the total of exercise done in the whole week. This contributed to building self-efficacy in the user's ability to do exercise.

## Persona

For the first iteration, no explicit persona was attached to the robot. Any persona assigned by the user was an interpretation of the user's perception of its physical appearance, voice, behavior, dialogue content, and presentation of the dialogue.

For the second iteration, an explicit attempt was made to increase consistency between the form of the robot, the behavior that it produced, and the expectations of the user. With an explicit, consistent design, the predictability of the robot's behavior increases, making it easier to interact with. It accounts for reducing potential dislike toward the robot as a result of inconsistency between verbal and nonverbal behavior, which ultimately predicts increased engagement between the user and the robot and therefore more data for analysis.

To enable consistency throughout the behavior of the agent, following Nass and Brave (2005), guidelines for the robot were established to define its set of behaviors or personality. The guidelines were based on users' perceptions of the robot to minimize mismatch between user expectations and subsequent perceived behavior, integrating, for example, the robot's physical form as a small rabbit. The set of guidelines were not intended to be revealed explicitly in the dialogue but provided a means to design the dialogue and behavior consistently. The explicitly designed persona allowed definition of the boundaries of the robot's behavior and knowledge and provided reasoning for that behavior. It defined when further explanations for behavior might be required and grounded the behavior in an explainable context.

The robot's persona was designed to be shy, polite, submissive, and self-deprecating, which is assumed to match the users' perception of a small animal and its role in passing on information when asked. It contributes to accounting for its behavior of sitting in the user's house and not always talking when the user is nearby. It was assumed that showing positive character traits such as being friendly, helpful, and likeable, and being keen and happy to fulfill the role that it has been given, would lead to a positive perception of the robot and contribute to increased engagement and therefore sufficient video data.

As part of the display of the robot's persona and behavior, the dialogue content specifically avoided any direct show of authority including advice, praise-giving, or reprimands. Observations from the first iteration of data collection suggested that any show of authority from the robot was not well received by the users, and to encourage more use of the system, this aspect was removed. As stated in the Department of Health's NHS Health Trainer's manual (Michie *et al.* 2008), giving advice may contribute to people feeling that they are not being listened to. Providing information to the

user without advice allows the user to decide what is relevant and is therefore more likely to make an impact on the modification or maintenance of the user's own behavior. The shyness of the robot contributes to the perception that the user is in control of the system and is dominant over the robot. The user is not derided for not achieving goals, just reminded that the information provided has been taken from the plan.

The most explicit change used to convey consistency between the robot's form and behavior was to develop a voice that is hypothesized to be more consistent with the robot's form. Used as a base voice, the female UK Text-to-Speech voice (Loquendo) was manipulated to have a higher pitch to match the smaller size of the robot, and, in combination with the high timbre, gave the voice a more comical and less human-like quality.

### **Attention Management**

During the first iteration, the robot's attention was managed via the dialogue interaction and the various sensor inputs. The robot would start an interaction if it had information about the participant performing some action, such as activating the key hook switch, and therefore was attending to the robot. On PIR sensor initiation of the dialogue, the robot is aware of the user's presence but it does not have any way of knowing whether the user is attending to it and therefore does not know how to respond and show that it is attending to the user.

For iteration 2, a more involved system was devised to separate what is said (the dialogue manager) and when something is said (the interaction manager) as introduced in Wallis (2010).

The SALT(E) (Sleeping, Alert, Listening, Talking, Engaged) interaction manager distinguishes among three states of the system:

- Sleeping: not seeing or hearing anything
- Alert: attending to the person
- Engaged: committed to a conversation, either Listening or Talking

As an example, the robot is in the sleep state when there is nobody present to activate any part of the system, indicated outwardly to the user by its ears being in a horizontal position (a. in Figure 4). When the user passes by or approaches the robot, activating the PIR sensor, the robot indicates that it is attending to the user by raising its ears to a vertical position (b. in Figure 4). If the robot has a conversation that it is ready to produce, its ears move to the engaged vertically raised position, with one ear slightly further forward than the other (c. in Figure 4), and the conversation is initiated. After the conversation, the system keeps the context of the recent

discussion and waits to see if the user wants to change the topic or input any other information to continue the conversation before moving back to the alert state and then after a period of time, the sleep state.

The PIR sensor is the primary means by which the system is moved from sleeping to alert, as the approach to activate the other switches is usually preceded by movement detection. The movement from alert to engaged is produced by either the system-initiated interactions, such as the morning dialogue or the key hook activated interactions, or by user-initiated activations when the user approaches the robot and shows it a face card.

Decisions about movement between these states and actions performed are controlled by time periods between events. There are four different types of pauses:

- Pause 1: indicates the end of a turn by the user and is the opportunity for the system to say something. This maps the movement between Listening and Talking.
- Pause 2: indicates that the system ought to say something, and, with nothing to say, it makes an encouragement. This maps the movement between Listening and Talking but inserts a conversation filler or offers help or advice on how to interact with the system.
- Pause 3: is the time after which the system drops the context of the conversation and moves from Engaged to Alert.
- Pause 4: is the time after the last PIR sensor activation by which the system recognizes that it does not need to attend. This maps the movement between Alert and Sleep.

The attention management system allows the robot to physically convey its presence and ability to interact without intruding on the user and demanding attention. The SALT(E) model explicitly addresses the issue that the robot is continuously switched on and present in a person's home rather than functioning as an object in a short-term experiment.

**TABLE 4** Showing the Number of Videos Gathered for Each Participant During Iteration 2 of Data Collection

Participant	N videos
1	28
2	14
3	5
4	25
5	36
<b>total</b>	<b>108</b>



## Data

The amount of data collected during iteration 2 is detailed in Table 4. Each interaction lasts between approximately ten seconds and up to approximately nine minutes, with an estimated majority of videos being up to two minutes in duration.

## ITERATION 3

P1, P2, P3, P4, and P5 were retained for iteration 3 and one new participant was recruited: P6 (see Table 1). The main changes made to the system are detailed below.

### Method of Interaction

The method of interaction remained the same, using RFID-tagged cards to input information. To assist with feedback of whether the tags had been read by the system, a noise was output, which indicated that the tag had been read. In addition to the extra feedback, the pause length where the robot waited for an input card response (see pauses 1 and 2 above in *Attention Management*) was increased to allow for potential problems in dealing with the complex nature of the interface and of the number of cards available.

### Dialogue Content

To physically demonstrate an empathetic response by the robot, during the self-evaluation interaction, questions were asked based on the feelings of the user and their rating between 1 and 5. If there was a response of 1 or 2, the robot moved its ears downward to the sleep position (a. in Figure 4) in combination with a verbal acknowledgement of disappointment with that rating. The ears subsequently returned to the engaged position (c. in Figure 4). The use and configuration of the ears for this task was informed by the study detailed in Eimler, Krämer, and von der Pütten (this issue).

The message card and dialogue was replaced with a “Recommendations” card, which attempted to introduce a more explicit interaction with the robot as an entity rather than as a conduit for information passed on from researchers. The robot stated that it had been looking for information for the user on social and community events and entertainment listings. This application has potential to be used to encourage social inclusion for older people or those with long-term conditions in order to improve overall well-being. The structure of the interaction was the same as that

of the message dialogue for iteration 2. The users were asked whether they would be interested in that recommendation, and they could rate it on the 1–5 scale or react with a face card. The aim was to provide personalized recommendations once a day for each participant.

In response to observations and data analysis from iteration 2 data, where the introduction the new card interface allowed more variability in the dialogue structure and therefore led to more issues regarding floor management for some participants, a card labelled “!”—termed the “frustration” card—was introduced. The card stopped the current interaction and sent the dialogue to the state where the user is in control and can either specify a topic to continue the conversation or leave the robot to go into the sleep state. It could be used at any time when the participant was frustrated with the robot or the direction of conversation and wanted to gain control of the interaction. Payr’s (this issue) analysis of iteration 2 data provides more detail on the issues that led to the attempt to resolve or influence floor management for iteration 3, and Heylen et al. (this issue) look at providing a model of floor management as an example of a key social skill, to implement into an agent architecture based on the output of data from iteration 2.

**Voice**

The voice used in this iteration was that used in the first iteration. This provided an opportunity for an assessment of attitudes by the users dependent on which voice was presented to them in their first contact with the robot.

**Data**

The amount of data collected during iteration 3 is detailed in Table 5. Each interaction lasts between approximately 10 seconds and up to

**TABLE 5** Showing The Number of Videos Gathered for Each Participant During Iteration 3 of Data Collection

Participant	N videos
1	6
2	25
3	0
4	12
5	30
6	39
<b>total</b>	<b>112</b>

approximately 5:20 minutes, with an estimated majority of videos being up to two minutes in duration.

**SUMMARY**

The SERA project robot setup was designed to provide three different sets of functions and behaviors with which to collect audio-visual data for analysis. The system was deployed three times to provide audio-visual data for analysis of the development of relationships over the 10-day trial period and also across the iterative deployment periods. The aim was not to empirically test hypotheses associated with the changes made for each iteration, but to alter the system sufficiently to gain new insights into the nature of the human-robot relationship that develops with the system alterations. This paper has described the systems deployed in the homes of real users for each of the three iterations. It has highlighted the strategy of meeting the data quantity needs of the researchers by making changes predicted to increase engagement for the participants and therefore to maintain or increase use of the system.

The total amount of data collected across all iterations is shown in Table 6. This methodology has produced a corpus of video data (with associated post-trial interviews) that provides a rich dataset for hypothesis generation, observation, and analysis on the nature of human-robot interaction in a real domestic environment. Analysis of data collected from a participant’s home has to address issues to account for the variable context and environment as opposed to the controlled conditions of a laboratory; see Wallis (this issue) for a discussion and proposed methodology and Krämer et al. (this issue), who highlight the need for a more qualitative approach to analysis of data of this type due to the idiosyncratic nature of participants’ interactions with the robot gathered during the SERA project. Collecting data in a domestic setting introduces challenges of recruiting participants who are to be filmed in their own homes for extended periods of time and ensuring their privacy. Managing the technology and equipment to maintain the functionality of all components for an extended

**TABLE 6** Showing the Total Number of Videos Comprising the SERA Corpus Across All Iterations of Data Collection

Iteration	N participants	N videos
1	3	52
2	5	108
3	6	112
	<b>total</b>	<b>272</b>

period of time in a new environment also provides a technical challenge in addition to designing and implementing the set of behaviors applied to each of the iterations. Overcoming these challenges, however, provides unique data, and through its analysis, insights into assistive and companion robots involving real potential end users of this technology and the real environments in which the robots could be used.

## NOTE

1. <http://www.sheffield50plus.co.uk>

## REFERENCES

- Bandura, A. 1977. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 84:191–215.
- Bandura, A. 1998. Health promotion from the perspective of social cognition theory. *Psychology and Health* 13:623–649.
- Bickmore, T., A. Gruber, and R. Picard. 2005. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education Counselling* 59 (1): 21–30.
- Brown, P., and S. C. Levinson. 1987. *Politeness*. Cambridge: Cambridge University Press.
- DelliFraine, J. L., and K. H. Dansky. 2008. Home-based telehealth: A review and meta-analysis. *Journal of Telemedicine and Telecare* 14:62–66.
- Eimler, S. C., N. C. Krämer, and A. von der Pütten. 2011. Determinants of acceptance and emotion attribution in confrontation with a robot rabbit. *Applied Artificial Intelligence* 25 (6): in press.
- Heylen, D., R. op den Akker, M. ter Maat, P. Petta, S. Rank, D. Reidsma, J. Zwiers. 2011. On the nature of engineering social artificial companions. *Applied Artificial Intelligence* 25 (6): in press.
- Klamer, T., and S. Ben Allouch. 2010a. Acceptance and use of a zoomorphic robot in a domestic setting. In *Proceedings of EMCSR*, 553–558. Vienna, Austria.
- Klamer, T., and S. Ben Allouch. 2010b. Acceptance and use of a social robot by elderly users in a domestic environment. In *Proceedings of Pervasive Health 2010*, 1–8. Munich, Germany.
- Klamer, T., S. Ben Allouch, and D. Heylen. 2010. Adventures of Harvey—Use, acceptance of, and relationship building with a social robot in a domestic environment. In *Proceedings of the 3rd International Conference on Human Robot Personal Relationships LNCS*, ed. F. J. Verbeek and M. H. Lamers, 59:74–85.
- Krämer, N., G. Bente, and J. Piesk. 2003. The ghost in the machine. The influence of Embodied Conversational Agents on user expectations and user behavior in a TV/VCR application. In *IMC Workshop 2003: Assistance, Mobility, Applications*, ed. G. Bieber and T. Kirste, 121–128. Rostock, Germany.
- Krämer, N. C., S. Eimler, A. von de Pütten, and S. Payr. 2011. Theory of companions: What can theoretical models contribute to applications and understanding of human-robot interaction? *Applied Artificial Intelligence* 25 (6): in press.
- Lewin, B., J. Pattenden, J. Ferguson, and H. Roberts. 2005. *The heart failure plan*. London: British Heart Foundation.
- Michie, S., N. Rumsey, A. Fussell, W. Hardeman, M. Johnston, S. Newman, and L. Yardley. 2008. Improving health: Changing behavior. In *Department of Health's NHS health trainer's manual*. London: Department of Health.
- Nass, C., and S. Brave. 2005. *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge, MA: MIT Press.
- Paré, G., M. Jaana, and C. Sicotte. 2007. Systematic review of home telemonitoring for chronic diseases: The evidence base. *Journal of the American Medical Informatics Association* 14:269–277.

- Payr, S. 2010. Analyzing video data from a field study. In *19th IEEE International Symposium in Robot and Human Interactive Communication (ROMAN)*, ed. C. A. Avizzano, E. Rufaldi, M. Carrozzino, M. Fontana, and M. Bergamasco, Viareggio, Italy.
- Prochaska, J., and W. Velicer. 1997. The transtheoretical model of behavior change. *American Journal of Health Promotion* 12:38–48.
- Wallis, P. 2010. A robot in the kitchen. In *Proceedings of ACL workshop on companionable dialogue systems*, 25–30. Uppsala, Sweden.
- Wallis, P., V. Maier, S. Creer, and S. Cunningham. 2010. Conversation in context: What should a robot companion say? In *Proceedings of EMCSR*, 547–552. Vienna, Austria.

Copyright of Applied Artificial Intelligence is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.